Introduction

Kalman filtering is a method for recursively updating an estimate μ of the state of a system by processing a succession of measurements Z. After each measurement, a new state estimate is produced by the *measurement* step. Z and μ do not necessarily have to have the same dimensionality; for example, μ could be an estimate of the two dimensional position of a target, which would be represented as a 2-by-1 matrix, whereas Z could be a bearing to the target (a scalar). Provision is made in the *movement* model for the possibility that the state of the system may change between measurements.

All of Kalman's computations can be thought of as manipulations of multivariate normal probability distributions. In fact, there are only two essential facts on which the whole structure is built, the first being associated with movement and the second with measurement. These facts are stated below using random variables X, V, and W, all independent of each other. Throughout these notes, symbols for matrices will be given in bold type. The *state* X is assumed to be multivariate normal with mean μ and covariance matrix $\mathbf{\Sigma}$, which we abbreviate $X \sim N(\mu, \mathbf{\Sigma})$. V is the measurement noise and W is the movement noise; we assume $V \sim N(\mu_V, \mathbf{R})$ and $W \sim N(\mu_W, \mathbf{Q})$. X and W are n-dimensional and V is m-dimensional. If $\boldsymbol{\phi}$ and $\mathbf{H}$ are appropriately dimensioned movement and measurement matrices, the two vital facts are (the superscript^t means transpose and $\mathbf{I}$ is the identity matrix):

1) If $X' = \boldsymbol{\phi}X + W$, then $X' \sim N(\mu', \mathbf{\Sigma}')$, where $\mu' = \boldsymbol{\phi}\mu + \mu_W$ and $\mathbf{\Sigma}' = \boldsymbol{\phi}\mathbf{\Sigma}\boldsymbol{\phi}^t + \mathbf{Q}$.

2) If $Z = \mathbf{H}X + V$, then, conditional on Z being given, $X \sim N(\hat{\mu}, \hat{\mathbf{\Sigma}})$, where $\hat{\mu} = \mu + \mathbf{K}(Z - \mu_V - \mathbf{H}\mu)$ and $\hat{\mathbf{\Sigma}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{\Sigma}$. The matrix $\mathbf{K}$ is called the Kalman gain, and is given by $\mathbf{K} = \mathbf{\Sigma}\mathbf{H}^t (\mathbf{H}\mathbf{\Sigma}\mathbf{H}^t + \mathbf{R})^{-1}$. In general, computation of K requires a matrix inverse.

Discussion of Fact 1:

The purpose of a KF is to keep track of the state of a system by making a sequence of measurements. It is permitted to have the state of the system change from X to X' between measurements, with the attractive feature of the movement model $X' = \boldsymbol{\phi}X + W$

being that normality is preserved. The formula for μ' should make intuitive sense, given the movement model. Recall that in the scalar case $\mathrm{Var}(\phi X) = \phi^2 \mathrm{Var}(X)$, so that the presence of both ϕ and ϕ^t in the formula for $\mathbf{\Sigma}'$ should come as no surprise. $\mathbf{Q}$ is additive because W is independent of X. ϕ and $\mathbf{Q}$ are n-by-n matrices, and μ_W is an n-vector. If the system state does not change between measurements (vacuous movement step), then $\phi = \mathbf{I}$, $\mu_W = 0$, and $\mathbf{Q} = 0$.

Discussion of Fact 2:

Suppose you were told that $X \sim N(-1, 4)$, and asked to guess X. You would probably guess "–1", the mean value. That answer can be justified as being optimal in almost any reasonable sense of the word. Suppose you were also told that $Z = -3.1$, and that Z was obtained by adding a "measurement noise" V to X, where $V \sim N(0, 2)$. Given this information, what would you guess for X? Since Z is smaller than –1, and since Z represents a reasonably accurate, unbiased ($\mu_V = 0$) measurement of X, your intuition would tell you to revise your estimate of X downward. In fact, since $2 < 4$, you would probably conclude that the best guess at X would be closer to –3.1 than to –1. The best way to make these intuitive considerations precise is to employ Bayes' Theorem, as is done in the appendix. The result of that application is Fact 2. In our scalar example, $\mathbf{\Sigma} = 4$, $\mathbf{H} = 1$, and $\mathbf{R} = 2$. It follows that $\mathbf{K} = 2/3$, $\hat{\mu} = (-1) + (2/3)(-3.1 + 1) = -2.4$, and $\hat{\mathbf{\Sigma}} = (1 - 2/3)\mathbf{\Sigma} = 4/3$. Not only does Bayes tell you to guess –2.4, but he tells you how accurate the guess is!

Fact 2 states that the best way to process the information in Z is to revise the "inputs" μ and $\mathbf{\Sigma}$ to the "outputs" $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$. The state of the system is still normal after the measurement is processed--the mean and covariance matrix have simply changed. The fact that normality is preserved is important, since $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$ may be the inputs to similar calculations in the future.

The simplicity of the way in which Kalman revises μ to $\hat{\mu}$ is also important. Note that $\mathbf{H}\mu + \mu_V$ is the mean or best guess of the measurement Z, so that $Z - \mu_V - \mathbf{H}\mu$ is the "shock" caused by the measurement*. If the measurement is not shocking, Kalman sets

* There is a different formula for shock if the KF is *extended* (see p. 12).

$\hat{\mu} = \mu$; otherwise, he makes a correction that is proportional to the shock. The matrix **K** is simply the proportionality constant. This method of revising μ to $\hat{\mu}$ is so simple and natural that the robustness of the procedure with respect to modeling errors should not be surprising.

Employment of Fact 2 requires one to know **H**, μ_V, and **R**. If Z has m components, then **H** is m-by-n, μ_V is m-by-1, and **R** is m-by-m.

Operation of the KF

There are two more required inputs: μ_0 and $\mathbf{\Sigma}_0$ are the initial values for μ and $\mathbf{\Sigma}$, respectively. Once μ and $\mathbf{\Sigma}$ are initialized, all calculations correspond to either movement or measurement, as exemplified in the diagrams in Figure 1, where the replacement symbol $\leftarrow$ makes it possible to dispense with the ' and ˆ notation used in stating Facts 1 and 2.

The $\leftarrow$ notation emphasizes that operation of a Kalman Filter can be thought of as a sequence of updates to μ and $\mathbf{\Sigma}$. Sufficient memory to store one copy of μ and one copy of $\mathbf{\Sigma}$ is all that is required when these updates are made by computer, as is usually the case. However, for tutorial purposes it is sometimes useful to let $(\mu_i(-), \mathbf{\Sigma}_i(-))$ be $(\mu, \mathbf{\Sigma})$ with all updates up to time t_i *except* for the measurement at time t_i, and to let $(\mu_i(+), \mathbf{\Sigma}_i(+))$ be similarly defined except that the update for the measurement at time t_i is included. Thus the measurement block of Figure 1 updates $(\mu_i(-), \mathbf{\Sigma}_i(-))$ to $(\mu_i(+), \mathbf{\Sigma}_i(+))$, while the movement block updates $(\mu_i(+), \mathbf{\Sigma}_i(+))$ to $(\mu_{i+1}(-), \mathbf{\Sigma}_{i+1}(-))$. Other matrices will also be subscripted for time in this expanded notation.

MOVEMENT
$\mu \leftarrow \phi\mu + \mu_w$
$\mathbf{\Sigma} \leftarrow \phi\mathbf{\Sigma}\phi^t + \mathbf{Q}$

MEASUREMENT (Z)
$\mathbf{K} \leftarrow \mathbf{\Sigma H}^t(\mathbf{H\Sigma H}^t + \mathbf{R})^{-1}$
$\mu \leftarrow \mu + \mathbf{K}(\mathbf{Z} - \mu_v - \mathbf{H}\mu)$
$\mathbf{\Sigma} \leftarrow (\mathbf{I} - \mathbf{KH})\mathbf{\Sigma}$

Figure 1. Showing the calculations corresponding to movement and measurement in a Kalman Filter.

Summary of Notation

$\boldsymbol{\phi}$ is the **movement matrix**, and is part of the description of how the state changes between measurements.

(μ_W, **Q**) is the mean and covariance of the **movement noise**. If you accidentally make **Q** too large, Kalman will be "high strung"; that is, his estimates will bounce around a lot because he takes the measurements too seriously because he make **K** too large. If you make **Q** too small, Kalman will be "lethargic" because he makes **K** too small.

H is the **measurement matrix** that describes how the measurement depends on the state.

(μ_V, **R**) is the mean and covariance of the **measurement noise**. Kalman's tendencies with respect to **R** are the opposite of those with **Q**.

(μ, $\boldsymbol{\Sigma}$) is the mean and covariance of the **state** of the system. μ can also be interpreted as Kalman's guess at the state of the system. The initial values (μ_0, $\boldsymbol{\Sigma}_0$) must be provided; after that, it is Kalman's job to continually update (μ, $\boldsymbol{\Sigma}$). It is sometimes useful to have a notation that distinguishes (μ, $\boldsymbol{\Sigma}$) before and after the i^{th} measurement is processed, in which case we will refer to ($\mu_i(-)$, $\boldsymbol{\Sigma}_i(-)$) as the "before" quantities and ($\mu_i(+)$, $\boldsymbol{\Sigma}_i(+)$) as the "after" quantities.

Z is the **measurement**.

K is the **Kalman gain**. This is used to update (μ, $\boldsymbol{\Sigma}$) by processing Z (see cover illustration).

Example 1

A target moves in a one-dimensional random walk, adding an increment to its position between observations that is normal with mean 1 mile and standard deviation 2 miles. In other words, $\boldsymbol{\phi} = 1$, $\mathbf{Q} = 4$ miles^2, and $\mu_W = 1$ mile. The state or position of the target is basically increasing with time, but the random component will cause occasional exceptions where the state decreases instead of increasing. We also assume $\mathbf{H} = 1$, $\mu_V = 0$, and $\mathbf{R} = 9$ miles^2, which is the same as saying that unbiased measurements of X are available that are accurate to within about $\sqrt{\mathbf{R}} = 3$ miles standard deviation. The initial guess at the target's position is $\mu_0 = 0$, $\boldsymbol{\Sigma}_0 = 10000$ miles^2; the large value for $\boldsymbol{\Sigma}_0$ indicates that whoever was forced to make the initial guess had basically no idea where the target was. Suppose the first three measurements are 84, 83, and 88, from which we might

conclude that the target's position is somewhere in the 80's even without Kalman's help. We take $(\mu_0, \boldsymbol{\Sigma}_0)$. to be (0, 10000), which essentially says that the initial guess is 0, but so inaccurate that it should not be trusted. Since a measurement is made before the target moves, we take $(\mu_1(-), \boldsymbol{\Sigma}_1(-))$ to be $(\mu_0, \boldsymbol{\Sigma}_0)$. The KF would use the measurement and movement blocks alternately, with the results for $(\mu, \boldsymbol{\Sigma})$ shown in Figure 2.

	Before Measurement *i*			**After Measurement *i***	
	$\mu_i(-)$	$\boldsymbol{\Sigma}_i(-)$	$\mathbf{K}_i$	$\mu_i(+)$	$\boldsymbol{\Sigma}_i(+)$
1	0	10000	1.00	84	9
2	85	13	.59	83.82	5.32
3	84.82	9.32	.51	86.44	4.58
4	87.44	8.58	.49	?	4.39
	?	⋮	⋮	?	⋮
∞	?	8.33	.48	?	4.33

Figure 2. Employment of a Kalman Filter to track a target.

Since the first Kalman gain is (10000/10009) ≈ 1, the filter forgets the initial guess entirely as soon as the first measurement is available; it guesses that the target's position is $\mu_1(+) = 84$, with the associated accuracy being the same as the accuracy of a measurement. It then adds 1 to $\mu_1(+)$ and 4 to $\boldsymbol{\Sigma}_1(+)$ to obtain $\mu_2(-)$ and $\boldsymbol{\Sigma}_2(-)$, reflecting the idea that the best guess of the target's position right before the second measurement is one unit larger than the best guess right after the first, but that $\mu_2(-)$ is a worse guess than $\mu_1(+)$ because of the unpredictable part of the target's motion. Turn the crank to make sure you see how the rest of the numbers are obtained. Letting $\mathbf{K}_i$ be the Kalman gain for the i^{th} measurement, you should get $\mathbf{K}_3 = 9.32/(9.32 + 9) = .51$, $\mu_3(+) = 84.82 + .51\ (88 - 84.82) = 86.44$, etc.

Note that $\mathbf{K}_4$ and $\boldsymbol{\Sigma}_4(+)$ can be computed even before the fourth measurement is made; in fact, the entire sequence of Kalman gains and variances is completely independent of the measurements. A close inspection of Figure 1 shows that this will always be the case; neither Z nor μ is ever used in computing $\boldsymbol{\Sigma}$ or $\mathbf{K}$. This could be an important feature in a situation where measurements had to be processed rapidly, since the Kalman gains could all be computed beforehand.

The movement block is a variance increasing operation, while the measurement block is a variance decreasing operation. This is evident in Figure 2. It sometimes

happens that the net result of these opposite forces is that the covariance matrices and Kalman gains approach steady state limits. Assuming they exist, the steady state limits $\boldsymbol{\Sigma}(-)$, $\mathbf{K}$, and $\boldsymbol{\Sigma}(+)$ must satisfy the equations

$$\begin{cases} \Sigma(-) = \phi\Sigma(+)\phi^t + \mathrm{Q} \\ \mathrm{K} = \Sigma(-)\mathrm{H}^t\left(\mathrm{H}\Sigma(-)\mathrm{H}^t + \mathrm{R}\right)^{-1} \\ \Sigma(+) = (\mathrm{I} - \mathrm{KH})\Sigma(-) \end{cases}$$

In our scalar example with $\boldsymbol{\Sigma} = \mathbf{H} = 1$, $\mathbf{R} = 9$, and $\mathbf{Q} = 4$, the only solution of these equations for which $\boldsymbol{\Sigma}(+)$ is positive is

$$\begin{aligned} \boldsymbol{\Sigma}(+) &= \left(\sqrt{\mathbf{Q}^2 + 4\mathbf{QR}} - \mathbf{Q}\right)/2 = 4.33 \\ \boldsymbol{\Sigma}(-) &= \boldsymbol{\Sigma}(+) + \mathbf{Q} = 8.33 \\ \mathbf{K} &= \boldsymbol{\Sigma}(-)/\left(\boldsymbol{\Sigma}(-) + \mathbf{R}\right) = .48 \end{aligned}$$

Comparison of these three numbers with the fourth row of Figure 2 shows that the steady state limit is approached rather quickly in this case. Note that the steady state accuracy of the filter's estimate right after a measurement ($\sqrt{\boldsymbol{\Sigma}(+)} = 2.08$ miles) is better than the accuracy associated with the latest measurement (3 miles). In fact, $\boldsymbol{\Sigma}(+)$ would be 0 if either $\mathbf{R}$ were 0 (that's obvious) *or* if $\mathbf{Q}$ were 0 (that's not obvious, but think about what happens when you can make lots of measurements of an unknown but *fixed* quantity). Note also that the steady state equations do not involve μ_0 or $\boldsymbol{\Sigma}_0$, which is a relief.

A particularly simple filter would use the steady state $\mathbf{K}$ at every stage and dispense with the covariance matrix calculations. Such a filter will typically behave poorly in the early stages unless the initial guess μ_0 happens to be close to the truth. But if only steady state performance is important, the simplicity of the technique could make it attractive. Try repeating the calculation of $\mu_4(-)$ in the example using $\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{K}_3 = .48$. Better yet, use $\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{K}_3 = .5$, since the basic point is that the filter will still do a good job even if the gain schedule isn't precisely as indicated in Figure 2.

Example 2

Suppose now that the target of Example 1 is actually quite massive, so that the erratic type of random walk motion postulated there is implausible. Suppose, instead, that the target can actually be thought of as having a velocity that changes by only a small amount (say .1 miles per hour) over the .5 hour time interval between position measurements. The "state" of the target is now (position, velocity)^t, a 2-by-1 vector. The movement model has

$$\boldsymbol{\phi} = \begin{bmatrix} 1 & .5\,\text{hours} \\ 0 & 1 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & .01\,\text{miles}^2/\text{hour}^2 \end{bmatrix} \quad \mu_W = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Assuming that the measurement is still of the target's position, **H** is now the 1-by-2 matrix **H** = [1 0], and **R** is as before. After initializing $\boldsymbol{\Sigma}_0$ and μ_0 (which are now 2-by-2 and 2-by-1 matrices), the movement and measurement blocks can be employed as before. If the new movement model is a better representation of how targets move than was the simple random walk of Example 1, then the position estimates (first component of μ) will be more accurate than they were before. This trick of improving accuracy by augmenting the state vector will come as no surprise to a reader familiar with Markov Chains. The main computational effort in implementing a Kalman Filter is in calculating **K**, where a matrix with the same dimensions as **R** must be inverted. Since the dimensions of this matrix are independent of the size of the state vector, the computational impact of such augmentation is fairly small.

The IOU Model and MTST

Let X_t be the velocity of a target at time t, and suppose that $X_{t+1} = X_t + W_t$ for $t \geq 1$, where W_1, W_2, … is a sequence of independent identically distributed normal random variables with mean 0 and variance q. This is a model of random walk. With q = .01 miles^2/hr^2, it would be the velocity model employed in Example 2. Since $\text{Var}(X_{t+1}) = \text{Var}(X_t) + q$, the sequence X_1, X_2, … has a progressively increasing variance. In the long run, according to this movement model, target speeds that exceed (say) 1000 miles/hr are not only possible but likely. Most real targets on earth cannot achieve such speeds, so this feature must be considered a modeling defect. However, a simple revision

can at least keep Var(X_t) within bounds, and thereby render such extreme deviations from 0 unlikely.

The revision is $X_{t+1} = cX_t + W_t$, where $0 \le c < 1$. Since $\mathrm{Var}(X_{t+1}) = c^2\mathrm{Var}(X_t) + q$, the shrinkage factor c will prevent Var(X_t) from growing large with t. In fact, Var(X_t) now has a limit s^2 as t approaches infinity, and this limit must satisfy the equation $s^2 = c^2s^2 + q$. If c and s are known, this equation can be solved for $q = s^2(1-c^2)$.

It is not hard to quantify or estimate s for a real target, since s is the target's root-mean-square velocity. However, it is also necessary to quantify c. To do so, consider forecasting X_{t+n} from a knowledge of X_t. By applying the movement model n times, it can be shown that $X_{t+n} = c^nX_t +$ (noise), where (noise) is a linear combination of W_t, W_{t+1}, …, W_{t+n-1}. The noise is 0 on the average. Now let $c = \exp(-\Delta/\tau)$, where Δ is the length of a time step, so that $c^n = \exp(-n\Delta/\tau)$. Since $n\Delta$ is the length of time over which the forecast is to be made, the parameter τ is a *relaxation time* for velocity. Thus the two target motion parameters that need to be quantified are τ and s. Given τ, s, and the time step length Δ, solve $c = \exp(-\Delta/\tau)$ and $q = s^2(1 - c^2)$ for the inputs to the movement model. For example, the East-West component of a ship's velocity might have $s = 5$ kt. and $\tau = 1$ hour. If $\Delta = .1$ hour (possibly because a measurement of the ship's position is made every 6 minutes), then $c = .905$ and $q = 4.532(\mathrm{kt})^2$. Figure 3 shows a Monte Carlo simulation of this movement model over an 8-hour period.

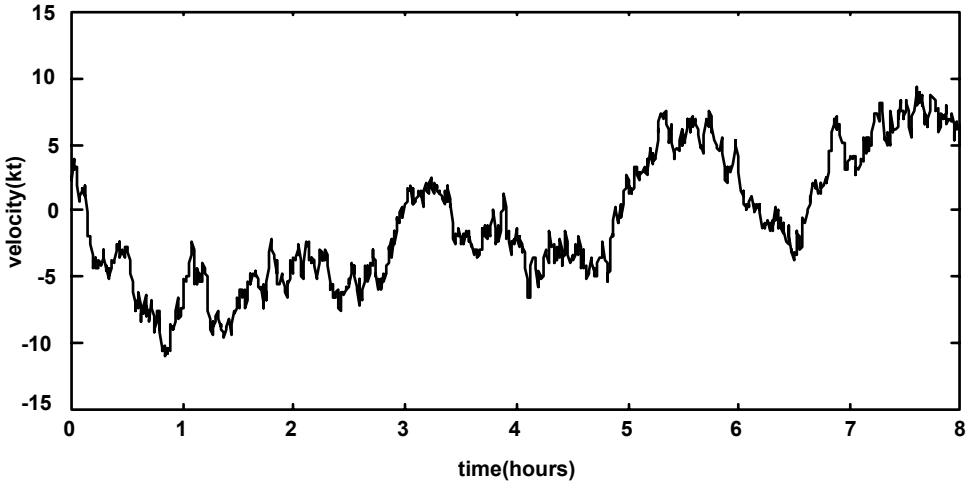

Figure 3. An O-U process fluctuating about 0.

The revised model described above is a discrete version of the Ornstein-Uhlenbeck (O-U) process, the most general normal Markovian stationary stochastic process with zero expectations. The O-U process is often used as a model of velocity because most velocities are zero on the average. Integrating velocity gives position, so the position of a target whose velocity is O-U is sometimes called IOU.

The U.S. Navy makes use of a Kalman Filter called the Maneuvering Target Statistical Tracker (MTST). MTST has a four component state vector, the first two being target location and the last two being target velocity. The two components of velocity are assumed to be independent O-U processes with the same (s, τ) parameters. The movement matrix is

$$\boldsymbol{\phi} = \begin{bmatrix} 1 & 0 & \delta & 0 \\ 0 & 1 & 0 & \delta \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & c \end{bmatrix}.$$

Parameter c has already been explained. Parameter δ is the multiplier for velocity used in updating position. Since velocity can fluctuate over the measurement interval Δ, and since the predictable part of velocity relaxes toward 0 with relaxation time τ, the velocity multiplier should be $\delta = \int_0^{\Delta} \exp(-\mu/\tau)du = \tau(1-c)$. δ is always smaller than Δ, but there is very little difference between the two when Δ is small compared to τ. The variance q appears in the **Q** matrix. The MTST measurement model depends on the type of measurement being made, but the movement model is always as described above.

The O-U process has been described as a model of velocity, but it can be adapted to any phenomenon that fluctuates around 0, with the fluctuation typically meaning "deviation from the intended quantity". For example, suppose that one component of the state vector is the target's course in degrees, and that the target's course fluctuates around 0 with time constant $\tau = 96$ seconds and standard deviation $s = 4$ degrees. Over an 8 second time interval, $c = \exp(-8/96) = .92$, and $q = s^2(1 - c^2) = 2.456$ in units of squared degrees. These are the parameters that would be used in the movement model for 8 seconds of movement.

Vacuous Movement Step

If the state does not change between measurements, then the movement step has no effect on either μ or $\mathbf{\Sigma}$, and operation of the filter amounts to processing a sequence of measurements on the same unknown state. In these circumstances it is usually best to keep track of the *inverse* $\mathbf{P} = \mathbf{\Sigma}^{-1}$ of the covariance matrix, rather than $\mathbf{\Sigma}$. $\mathbf{P}$ is often called the *precision matrix*. In terms of $\mathbf{P}$, the measurement step (see the Appendix) is

MEASUREMENT
$\mathbf{P} \leftarrow \mathbf{P} + \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H}$
$\mathbf{K} \leftarrow \mathbf{P}^{-1}\ \mathbf{H}^t\ \mathbf{R}^{-1}$
$\mu \leftarrow \mu + \boldsymbol{K}(Z - \mu_V - \mathbf{H}\mu)$

Note the simplicity of the update for $\mathbf{P}$; the formula makes it obvious that any measurement must literally add to the precision with which the state is known. Note also that $\mathbf{K}$ is calculated *after* $\mathbf{P}$ is updated.

Since $\mathbf{P}$ can only grow as more and more measurements are processed, the gain matrix $\mathbf{K}$ will eventually, in most cases, approach 0. This phenomenon sometimes referred to as "covariance collapse." There is nothing wrong with this if the state is indeed known to be unchanging, but if the state *should* change, a collapsed filter will not be able to track the change. Covariance collapse can be prevented by making $\mathbf{Q}$ nonzero, but in that case the movement step is no longer vacuous.

In a sequence of measurements, let $\mathbf{P}_i$ be the reciprocal of $\mathbf{\Sigma}_i(+)$ (alternatively the reciprocal of $\mathbf{\Sigma}_{i+1}(-)$, since the two covariance matrices are identical) and $\mu_i = \mu_i(+)$. The ± modifiers are no longer needed when there is no movement. μ_i is simply the best estimate of the state after i measurements have been processed. Taking $\mathbf{P}_0$ to be 0 is common, since doing so has the effect of giving no weight to *a priori* judgments about the state.

Example 3

Consider the linear regression of Y on X, where $(x_i, y_i) = (2,7), (0,2), (5,14)$ for $i = 1, 2, 3$. The usual approach is to find the best linear fit of the form $Y = aX + b$, where a and b are chosen to minimize the least squares expression $\sum_{i=1}^{3}(y_i - ax_i - b)^2$. The solution is $(a, b) = (2.395,2.079)$, obtained using the regression function on a hand calculator. This result can also be obtained using a Kalman filter where $(a, b)^t$ is regarded as the unknown (but unchanging, so there is no movement) state vector, with μ_0 and $\mathbf{P}_0$ both taken to be 0. The three measurements can be thought of as one super measurement $Z = (y_1, y_2, y_3)^t$, with $\mathbf{H} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{bmatrix}$ and $\mathbf{R} = r\mathbf{I}$, r being a scalar representing the variance associated with each observation and $\mathbf{I}$ a 3-by-3 identity matrix. Carrying out the measurement step, we find that

$$\mathbf{P}_1 = 0 + \frac{1}{r}\mathbf{H}^t\mathbf{H} = \frac{1}{r}\begin{bmatrix} \sum_{i=1}^{3} x_i^2 & \sum_{i=1}^{3} x_i \\ \sum_{i=1}^{3} x_i & 3 \end{bmatrix} = \frac{1}{r}\begin{bmatrix} 29 & 7 \\ 7 & 3 \end{bmatrix},$$

$$\mathbf{K}_1 = \frac{r}{38}\begin{bmatrix} 3 & -7 \\ -7 & 29 \end{bmatrix}\mathbf{H}^t\left(\frac{1}{r}\right) = \frac{1}{38}\begin{bmatrix} -1 & -7 & 8 \\ 15 & 29 & -6 \end{bmatrix}, \text{ and}$$

$$\mu_1 = 0 + \mathbf{K}(Z - 0) = \mathbf{K}\begin{bmatrix} 7 \\ 2 \\ 14 \end{bmatrix} = \frac{1}{38}\begin{bmatrix} 91 \\ 79 \end{bmatrix} = \begin{bmatrix} 2.395 \\ 2.079 \end{bmatrix},$$

the same result obtained by linear regression.

The point here is not that Kalman filtering is an easier way to do linear regression (far from it), but that the Kalman filter behaves as it should in a familiar situation. Note that the unknown variance r doesn't enter into calculating μ_1 because it cancels in the computation of $\mathbf{K}_1$. μ_0 was taken to be 0 for convenience in calculating μ_1, but any other estimate would have produced exactly the same result. Try it. Another good exercise would be to show that the same result would be obtained even if the three measurements were not processed all at once. Try processing any two of them in a batch, and then

sequentially processing the third. Regardless of which two you select to process first, you should get the same final result. If you attempt to process only one measurement first, you will find that the first calculated precision matrix doesn't have an inverse. This is Kalman's way of protesting the task you have set him, namely the task of estimating two numbers from only one measurement when the prior estimate is valueless.

In linear regression problems where the index i represents time, analysts sometimes worry about the parameters a and b changing slightly as time goes by. Intuitively, recent measurements should play a stronger role than old ones in estimating the state. This is exactly what would happen in a Kalman filter with a non-vacuous movement step where **Q** is positive, rather than 0.

Extended kalman filters (EKF'S)

If the measurement is a nonlinear function of the state variables, then the matrix **H** must be obtained by *linearizing* the nonlinear function. Formally, if $Z = f(X) + V$, then $\mathbf{H} = df(X)/dX|_{X=\mu}$ is the matrix of first partial derivatives (Jacobian). Except for the fact that **H** now depends on μ and that the shock is now $Z - f(\mu) - \mu_V$, the measurement step can be employed as before. The matrix **H** is used in calculating the Kalman gain, but it is not used in calculating the shock (the nonlinear function itself is used rather than the linear approximation). Similarly, if the movement model $X' = g(X) + W$ includes a nonlinear function g, then $\phi = dg(X)/dX|_{X=\mu}$ is the n-by-n Jacobian of g, and the movement step can be employed as before except that $\mu \leftarrow g(\mu) + \mu_W$. Note that the matrix ϕ is not used in updating μ; its use instead is in updating Σ. In either case, the result is called an EKF.

Example 4. Triangulation

The problem of estimating the position of a stationary target from several inaccurate bearing measurements can be solved by employing an EKF with a vacuous movement step. Let the state be $(x, y)^t$, with polar coordinates relative to an observer (r,θ) as in Figure 4. Observer i is located at (x_i, y_i), i=1,2,3.

Consider the measurement $Z_1 = \theta_1 + V_1$, where $\theta_1 = \arctan((y-y_1)/(x-x_1))$ is a nonlinear function of the state. Since $d\theta_1/dy = \cos(\theta_1)/r_1$ and $d\theta_1/dx = -\sin(\theta_1)/r_1$, **H** (for the first measurement alone) is $[-\sin\theta_1 \cos\theta_1]/r_1$, a 1-by-2 matrix. Since θ_1 and r_1 depend

on the unknown state, **H** must in practice be evaluated by inserting the latest estimates of θ_1 and r_1. Assuming $\mathbf{P}_0 = 0$, the precision matrix after one observation is

$$\mathbf{P}_1 = \mathbf{H}_1^t \mathbf{R}_1^{-1} \mathbf{H}_1 = \mathbf{R}_1^{-1} \begin{bmatrix} \sin^2\theta_1 & -\sin\theta_1\cos\theta_1 \\ -\sin\theta_1\cos\theta_1 & \cos^2\theta_1 \end{bmatrix} r_1^{-2},$$

where $\mathbf{R}_1$ is the variance of the angular measurement in radians^2. The product $\mathbf{R}_i r_i^2$ is the variance of the i^{th} measurement expressed as a distance in the vicinity of the target. Let $d_i \equiv \mathbf{R}_i r_i^2$. Then after n measurements,

$$\mathbf{P}_n = \begin{bmatrix} \sum_{i=1}^{n} \frac{\sin^2\theta_i}{d_i} & -\sum_{i=1}^{n} \frac{\sin\theta_i\cos\theta_i}{d_i} \\ -\sum_{i=1}^{n} \frac{\sin\theta_i\cos\theta_i}{d_i} & \sum_{i=1}^{n} \frac{\cos^2\theta_i}{d_i} \end{bmatrix}.$$

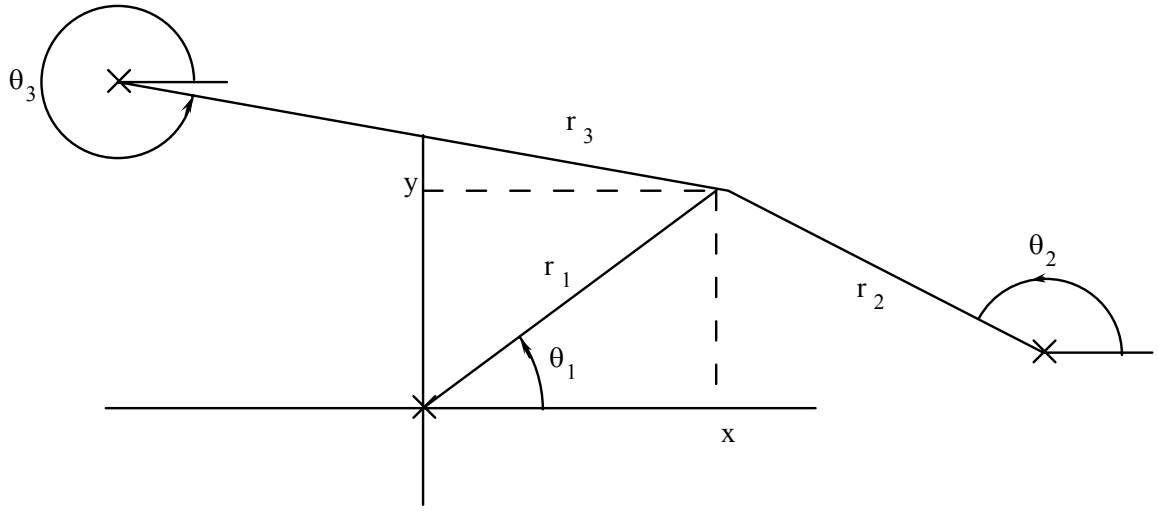

Figure 4. Three stations measure the bearing to a target.

We have not tracked through the state updates, but the shape and orientation of the bivariate normal distribution of the position of the target is already implied in **P**. The standard way of representing this distribution is to show the "two sigma ellipse," an equiprobability contour that contains the state with probability 1–exp(–2) = .865. This ellipse has its major axis at inclination I, a major diameter of length $4s_1$ and a minor diameter of length $4s_2$, where the inclination I and the two standard deviations s_1 and

s_2 can be determined from **P**. For the sake of completeness, we record the formulas, letting $\mathbf{P}^{-1} = \Sigma = \begin{bmatrix} a & h \\ h & b \end{bmatrix}$:

$$s_1^2 = \left[\left(\frac{a+b}{2} \right) + \sqrt{\left(\frac{a-b}{2} \right)^2 + h^2} \right]$$

$$s_2^2 = \left[\left(\frac{a+b}{2} \right) - \sqrt{\left(\frac{a-b}{2} \right)^2 + h^2} \right]$$

$$I = \begin{cases} \frac{1}{2} \tan^{-1}\left(\frac{2h}{a-b} \right) & \text{if } a > \text{b} \\ \frac{1}{2} \tan^{-1}\left(\frac{2h}{a-b} \right) + \frac{\pi}{2} & \text{if } a < \text{b} \\ \frac{\pi}{4} \operatorname{sign}(h) & \text{if } a = \text{b} \end{cases}$$

Example 5

Suppose $X = (R, \theta, U, \Psi)^t$, with the four components of state being range, bearing, speed and course; i.e., the state is position and velocity in polar coordinates. Active radars and sonars can often measure a Doppler shift in the transmitted frequency that is equivalent to observing a noisy version of $U\cos(\theta - \psi)$, the rate at which the range is changing. Suppose that noisy measurements of range and bearing are also available, so that each pulse results in a three-dimensional measurement:

$Z \leftarrow (R, \theta, U\cos(\theta - \psi))^t + V$. Since the cosine is a nonlinear function, an extended Kalman filter is required, with (letting $S = \sin(\theta - \Psi)$ and $C = \cos(\theta - \Psi)$)

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -US & C & US \end{bmatrix}.$$

Virtually any other measurement could be processed in the same manner as range rate. It is not required that each measurement be a direct measurement of a state variable, but only that each component should depend in some known manner on the state variables.

Since the quantities *U, S*, and *C* depend on the best current estimate of the state vector, the filter is an EKF. **H** will be a different matrix with each employment of the measurement step. Similar comments hold for the matrix ϕ, although the precise specification of ϕ depends on the movement model.

Since it is derived via Bayes Theorem, a (non-extended) Kalman filter makes optimal estimates in almost any reasonable sense of the word, provided it is applied in circumstances where the assumptions made in deriving it are valid. This is not true of the Extended Kalman Filter. Since the matrices **H** and/or ϕ depend on current state estimates and are used to obtain revised state estimates, there is a potential for bad estimates to get worse, and complete loss of track is possible. The design of practical filters in such circumstances is an art wherein one attempts to retain the behavior of the Extended Kalman Filter when things are going well, while simultaneously being able to recognize and correct for an incipient loss of track. Dimensionless shock plays a role in the latter.

Dimensionless Shock

The shock S_i used in making the measurement update at time i in an EKF is $Z_i - f(\mu_i(-)) - \mu_V$, the difference between what is actually measured (Z_i) and the best prediction of Z_i based on all history previous to the i^{th} measurement ($f(\mu_i(-)) + \mu_V$). One symptom of being out of control is that S_i is larger than can reasonably be explained by randomness. A useful measure of tracking quality can be built on this observation, provided a scale can be found on which S_i can be judged to be "unusually large."

The measurement model is that $Z_i = f(X_i) + V_i$. Approximating $f(X_i)$ by $f(\mu_i(-)) + \mathbf{H}_i(X_i - \mu_i(-))$ leads to the approximation $S_i \approx \mathbf{H}_i(X_i - \mu_i(-)) + (V_i - \mu_V)$, a linear combination of the two independent random variables X_i and V_i. The approximate expected shock is 0, and the approximate variance of S_i is $\mathbf{H}_i\mathbf{\Sigma}_i(-)\mathbf{H}^t + \mathbf{R}$. This variance is just the denominator of the Kalman gain computation, and it is the desired means for judging when S_i is "too large." Let the dimensionless shock be

$$DS_i \equiv S_i^t\left(\mathbf{H}_i\mathbf{\Sigma}_i(-)\mathbf{H}_i^t + \mathbf{R}\right)^{-1}S_i.$$

If S_i has m components, DS_i should be a scalar random variable that has a Chi-square distribution with m degrees of freedom. Thus if DS_i becomes large *compared to m,* the

likely explanation is that the filter has lost track. Note that the required matrix inverse is the same as is required to calculate $\mathbf{K}_i$, so very little additional effort is required to also calculate DS_i.

In addition to its use in recognizing when an EKF has lost track, the dimensionless shock also has a use in associating data with targets. Suppose that several targets are being tracked simultaneously, and that any of them might have caused a particular measurement. Other things being equal, the best target to associate with the measurement will be the one for which the dimensionless shock is smallest. Alternatively, iff *DS* is large for all targets, then the measurement might have been caused by some previously undiscovered target. Associating data with targets is an important part of *data fusion*. See Bar-Shalom and Fortmann [7] for an in-depth treatment.

Further Reading

Kalman filtering was invented by and for the most part has been used by electrical engineers, even though there is nothing even faintly electrical about it. This explains the name, and also the fact that a great deal of the literature is in the IEEE Transactions series. Kailath [3] provides an excellent historically based review with 390 (!) references. He traces the underlying ideas back to the work of Kolmogorov, Krein, and Weiner in the 1940s, and even beyond. Kalman's essential contribution was to recognize that the required computations can be done recursively; the seminal paper is [4].

References [1, 2, 5] are textbooks. The movement and measurement steps can be justified as being optimal in a least squares sense even when the assumptions of normality made here are abandoned, and that approach is in fact the one more commonly pursued. There is also a continuous time KF that is obtainable as a limiting form of the discrete filter; it involves ordinary differential equations for μ and $\mathbf{\Sigma}$ instead of the replacement ($\leftarrow$) operation of Figure 1.

The most important application of KFs so far has been to tracking problems where the state is a more or less elaborate description of the position of something. There are several examples in Gelb [1] and Titus [6].

Appendix

Theorem: Let $Z = \mathbf{H}X + V$, where X and V are independent, multivariate normal random variables for which $E(V) = 0$, $\mathrm{Cov}(V) = \mathbf{R}$, $E(X) = \mu$, and $\mathrm{Cov}(X) = \mathbf{\Sigma}$. Then, conditional on Z being given, X is multivariate normal with mean $\hat{\mu} = E(X|Z) = \mu + \mathbf{K}(Z - \mathbf{H}\mu)$ and covariance $\hat{\mathbf{\Sigma}} = \mathrm{Cov}(X|Z) = (\mathbf{I} - \mathbf{KH})\mathbf{\Sigma}$, where $\mathbf{K} = \mathbf{\Sigma}\mathbf{H}^t(\mathbf{H}\mathbf{\Sigma}\mathbf{H}^t + \mathbf{R})^{-1}$.

Proof: The joint density function of Z and X is (const.) exp $(-q/2)$, where $q = (x-\mu)^t \hat{\mathbf{\Sigma}}^{-1}(x-\mu) + (z - \mathbf{H}x)^t\mathbf{R}^{-1}(z - \mathbf{H}x)$. By simple expansion by terms, one can show that $q = (x-\hat{\mu})^t \hat{\mathbf{\Sigma}}^{-1}(x-\hat{\mu}) +$ (const.), where (const.) does not depend on x, $\hat{\mathbf{\Sigma}}^{-1} = \mathbf{\Sigma}^{-1} + \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H}$, and $\hat{\mu} = \mu + \hat{\mathbf{\Sigma}}\,\mathbf{H}^t\mathbf{R}^{-1}(z - \mathbf{H}\mu)$. This verifies that $E(X|Z)$ and $\mathrm{Cov}(X|Z)$ are $\hat{\mu}$ and $\hat{\mathbf{\Sigma}}$, respectively, but in most cases formulas that do not require inversion of $\mathbf{\Sigma}$ will be more convenient. To obtain such, we use the fact that $(\mathbf{I} + \mathbf{UH})^{-1}\mathbf{U} = \mathbf{U}(\mathbf{HU} + \mathbf{I})^{-1}$ for any matrix $\mathbf{U}$—note that the formula connects two matrix inversions of different dimension. In the following, we will define $\mathbf{U} = \hat{\mathbf{\Sigma}}\,\mathbf{H}^t\mathbf{R}^{-1}$. We have

$$\begin{aligned}
\hat{\mathbf{\Sigma}}\,\mathbf{H}^t\mathbf{R}^{-1} &= (\mathbf{\Sigma}^{-1} + \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1}\,\mathbf{H}^t\mathbf{R}^{-1} \\
&= (\mathbf{I} + \mathbf{UH})^{-1}\,\mathbf{U} = \mathbf{U}(\mathbf{HU} + \mathbf{I})^{-1} \\
&= \mathbf{\Sigma}\mathbf{H}^t(\mathbf{H}\mathbf{\Sigma}\mathbf{H}^t + \mathbf{R})^{-1} \\
&\equiv \mathbf{K},
\end{aligned}$$

thus showing that the formula given for $\hat{\mu}$ in the statement of the theorem is correct. To show that the formula for $\hat{\mathbf{\Sigma}}$ is correct, we will use the fact that $(\mathbf{I} + \mathbf{UH})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{UH})^{-1}\,\mathbf{UH}$. Since we showed above that $(\mathbf{I} + \mathbf{UH})^{-1}\,\mathbf{U} = \mathbf{K}$, it follows by substitution that $(\mathbf{I} + \mathbf{UH})^{-1} = \mathbf{I} - \mathbf{KH}$. We therefore have $\hat{\mathbf{\Sigma}} = (\mathbf{\Sigma}^{-1} + \mathbf{H}^t\mathbf{R}^{-1}\mathbf{H})^{-1} = (\mathbf{I} + \mathbf{UH})^{-1}\mathbf{\Sigma} = (\mathbf{I} - \mathbf{KH})\mathbf{\Sigma}$, showing that the expression for $\hat{\mathbf{\Sigma}}$ is also correct.

References

[1] Gelb, A. (1974), *Applied Optimal Estimation*, MIT Press, Cambridge, MA.

[2] Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, NY.

[3] Kailath, T. (1984), "A view of three decades of linear filtering theory," *IEEE Trans. Info. Theory*, IT-20, 146-180.

[4] Kalman, R. E. (1960), "A new approach to linear filtering and prediction problems," *Trans. ASME--J. Basic Engineer*, 82D, 34-45.

[5] Meditch, J. S. (1969), *Stochastic Optimal Linear Estimation and Control*, New York: McGraw-Hill.

[6] Titus, H. (1977), *Advances in Passive Target Tracking*, NPS report NPS62Ts-77071.

[7] Bar-Shalom, Y. and Li, X. (1995) *Multitarget-Multisensor Tracking: Principles and Techniques* , Academic Press